



TECHNICAL REPORT NO. TR-2010-07

**AN ACCUMULATE-TOWARD-THE-MODE
APPROACH TO CONFIDENCE INTERVALS AND
HYPOTHESIS TESTING WITH APPLICATIONS TO
BINOMIALLY DISTRIBUTED DATA**

February 2010

Approved for public release; distribution is unlimited.

**US ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ABERDEEN PROVING GROUND, MARYLAND 21005-5071**

DISCLAIMER

The findings in this report are not to be construed as an official Department of the Army position unless so specified by other official documentation.

WARNING

Information and data contained in this document are based on the input available at the time of preparation.

TRADE NAMES

The use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial hardware or software. The report may not be cited for purposes of advertisement.

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE Technical Report		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE An Accumulate-Toward-the-Mode Approach to Confidence Intervals and Hypothesis Tests With Applications to Binomially Distributed Data				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Dr. Dwayne W. Nuzman				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Director US Army Materiel Systems Analysis Activity 392 Hopkins Road Aberdeen Proving Ground, MD 21005-5071				8. PERFORMING ORGANIZATION REPORT NUMBER TR-2010-07	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					

20100301277**13. SUPPLEMENTARY NOTES****14. ABSTRACT**

Analyses of binomially distributed data usually depend on the normal approximation to the binomial or, for small sample sizes, the binomial cumulative distribution function. The first approach is only good for an unspecified "sufficiently large" sample size and does not reflect the asymmetry of the binomial distribution for probabilities other than 0.5. The second approach leads to suboptimal designs. Neither approach extends easily to handle multivariate binomial data. This paper describes a new accumulate toward the mode approach to hypothesis tests and confidence intervals and the application of this approach to binomially distributed data. The extension of this approach to multivariate data is also covered.

15. SUBJECT TERMS**16. SECURITY CLASSIFICATION OF:**a. REPORT
UNCLASSIFIEDb. ABSTRACT
UNCLASSIFIEDc. THIS PAGE
UNCLASSIFIED**17. LIMITATION
OF
ABSTRACT**
SAME AS
REPORT**18. NUMBER
OF PAGES**

33**19a. NAME OF RESPONSIBLE PERSON**

19b. TELEPHONE NUMBER (include area code)

Standard Form 298 (Rev.

8-98)

Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK.

CONTENTS

	Page
ACKNOWLEDGEMENTS	vi
LIST OF ACRONYMS	vii
1. Introduction.....	1
2. Hypothesis tests and confidence intervals	2
3. Choosing a rejection region	3
3.1 Symmetric distributions.....	3
3.2 Non symmetric distributions.....	3
3.3 The accumulate-toward-the-mode approach.....	4
4. Binomial distributions.....	6
4.1 Confidence intervals	6
4.1.1 The normal approximation.....	6
4.1.2 The binomial CDF	7
4.1.3 The binomial AMDF.....	7
4.2 Confidence regions for multivariate distributions	9
4.2.1 The AMDF for a bivariate binomial distribution.....	9
5. Examples from the helmet study.....	12
5.1 The original question	12
5.2 The amended question	13
5.3 Comparing methods	15
6. Observations and conclusions	18
REFERENCES	19
APPENDIXES	
A – MATLAB® Implementation	A-1
B – DISTRIBUTION LIST	B-1

LIST OF FIGURES

Figure No.	Title	Page
1	Rejection region for a symmetric distribution.....	3
2	Rejection regions for a non symmetric distribution.....	4
3	The accumulate-toward-the-mode distribution function.....	5
4	The binomial AMDF.....	7
5	Family of binomial AMDF functions for a sample size of 10.....	8
6	An example of the bivariate binomial AMDF.....	10
7	Confidence regions for bivariate binomial distributions.....	11
8	Maximum values of AMDF(X,Y) on the line $p_x = p_y$ for various absolute differences ($ X-Y $).....	13
9	AMDF(Z) for various outcomes.....	14
10	Comparison of Binomial and Normal 95% confidence intervals for the helmet study.....	15
11	Actual confidence levels.....	16
12	Actual confidence levels.....	16

LIST OF TABLES

Table No.	Title	Page
1	Iterative Procedure for choosing the rejection region.....	4

ACKNOWLEDGEMENTS

The U.S. Army Materiel Systems Analysis Activity (AMSAA) recognizes the following individuals for their contribution to this report.

The author is:

Dwayne Nuzman

Administrative Support:

Brenda Platt, Strategic and Operational Support Division (SOSD), AMSAA

LIST OF ACRONYMS

AMDF	- Accumulate-toward-the-Mode Distribution Function
AMSAA	- U.S. Army Materiel Systems Analysis Activity
CDF	- Cumulative Distribution Function
ICEM	- Integrated Casualty Estimation Model
PDF	- Probability Density Function

AN ACCUMULATE-TOWARD-THE-MODE APPROACH TO CONFIDENCE INTERVALS AND HYPOTHESIS TESTING WITH APPLICATIONS TO BINOMIALLY DISTRIBUTED DATA

1. Introduction.

Analyses of binomially distributed data usually depend on the normal approximation to the binomial or, for small sample sizes, the binomial cumulative distribution function. The first approach is only good for an unspecified “sufficiently large” sample size and does not reflect the asymmetry of the binomial distribution for probabilities other than 0.5. The second approach leads to suboptimal designs. Neither approach extends easily to handle multivariate binomial data.

This paper describes a new accumulate-toward-the-mode approach to hypothesis tests and confidence intervals and the application of this approach to binomially distributed data. The usual cumulative distribution function, which accumulates probability from left to right, is replaced by an accumulate-toward-the-mode distribution function which accumulates probability from areas of lower probability to areas of higher probability. This approach, when applied to asymmetrically distributed data, leads to more powerful hypothesis tests and more accurate interval estimates. In addition, this approach easily extends to analysis of multivariate distributions, providing decision makers with better information on the relative performance of alternatives.

Section 2 provides some background on the relationship between hypothesis tests and confidence intervals; gives examples with both symmetric and non symmetric distributions; and demonstrates how the usual procedures, when applied to non symmetric distributions, lead to suboptimal designs.

Section 3 introduces an alternative to the usual cumulative distribution function (CDF) for use in hypothesis testing and finding confidence intervals. For reasons that will become apparent, this new function is called the accumulate-toward-the-mode distribution function (AMDF). Its application to hypothesis testing and finding confidence intervals leads to optimal designs for both symmetric and non symmetric distributions.

In Section 4, the AMDF is developed for both the binomial and the bivariate binomial distributions. Section 5 contains examples of hypothesis tests and confidence intervals. This section also provides a comparison between confidence intervals based on the AMDF, and those typically obtained using the normal approximation. Section 6 gives some observations and conclusions.

2. Hypothesis tests and confidence intervals.

Statistics texts do not always point out the relationship between hypothesis testing and confidence intervals. This is unfortunate, since understanding this relationship can lead to better understanding of both. Suppose X is a random sample and p is a parameter to be estimated from X . Consider a collection of hypothesis tests, all with significance level α , and all with a null hypothesis of the form $p = p_0$, where p_0 may be any of the possible values of the parameter p . The set of p_0 values for which the null hypothesis cannot be rejected is a $100 \cdot (1 - \alpha)\%$ confidence interval for p .

There is some flexibility in the choice of rejection region for these hypothesis tests. This choice can affect the power of the tests and the size of the confidence interval. Under the null hypothesis (i.e., if the null hypothesis is true), the probability that the test statistic, T , will be in the rejection region must be α . That is the meaning of significance level. However there can be many sets with that probability; hence the flexibility in choice of rejection region. Ideally, one would like a rejection which meets the following conditions. Under the null hypothesis, the probability density of T is relatively low in the rejection region and relatively high in the acceptance region (i.e., outside the rejection region). Conversely, if the null hypothesis is not true, we would prefer the opposite conditions. That is, the probability density of T should be relatively high in the rejection region and relatively low in the acceptance region.

3. Choosing a rejection region.

3.1 Symmetric distributions. Figure 1 illustrates the choice of rejection region that best meets the above criteria when the distribution is symmetric. The curve in the upper subplot is a probability density function (PDF), representing the PDF of the test statistic under the true hypothesis. The curve in the lower subplot is the corresponding CDF. Any horizontal line, such as those in the upper subplot, could be used to define the rejection region: namely the area where the curve lies at or below the horizontal line. This choice automatically meets the criteria that the PDF of T be relatively low in the rejection region and relatively high in the acceptance region.

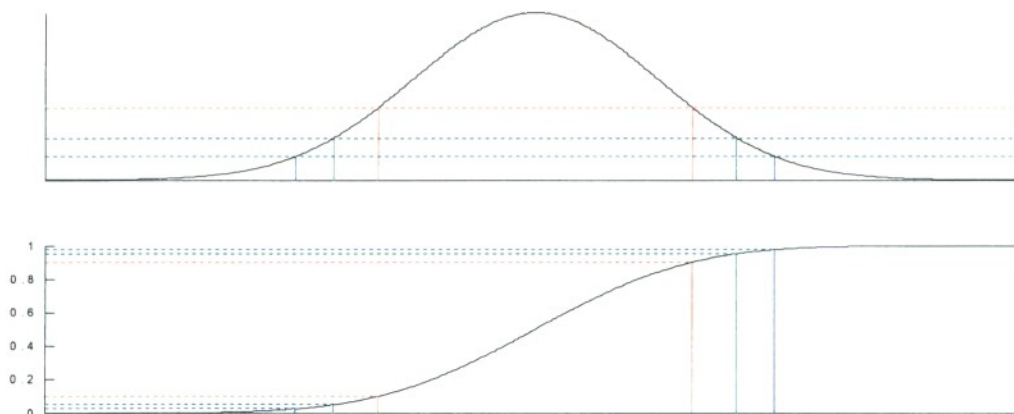


Figure 1. Rejection region for a symmetric distribution.

How well this choice meets the design criteria for the case where the null hypothesis is not true is more difficult to assess, since, in that case, the true distribution is unknown. That issue is generally addressed by increasing the sample size. That is a topic will not be dealt with in this paper.

At this point, the problem is how to select the line that will make the probability of rejection equal to α . That is where the CDF comes in.

The rejection region we have defined has two tails. Because of the symmetry of the distribution, we know that the probabilities of these two tails are both equal to $\alpha/2$. So, horizontal lines in the lower subplot at the levels $\alpha/2$ and $1-\alpha/2$, meet the CDF curve at the two cutoff points for the rejection region. The blue, green, and red lines in the figure correspond to significance levels of 0.05, 0.1, and 0.2 respectively.

Although the PDF function was used to illustrate the choice of rejection region, it was not needed to determine the threshold values. Only the CDF was required. This is why most statistical tables give values of the CDF but not the PDF.

3.2 Non symmetric distributions. The most commonly used approach for finding the rejection region when the distribution of T is non symmetric is to use the procedure just described for symmetric distributions. This is easily done with existing tables of the CDF function.

However, when the distribution is not symmetric, this approach fails to meet our criteria for selection of the rejection region. The upper subplot of Figure 2 gives examples of rejection regions selected in this way. As above, rejection regions for significance levels of 0.05, 0.1, and 0.2 are shown. In each case, there are densities in the acceptance region that are lower than densities in the rejection region.

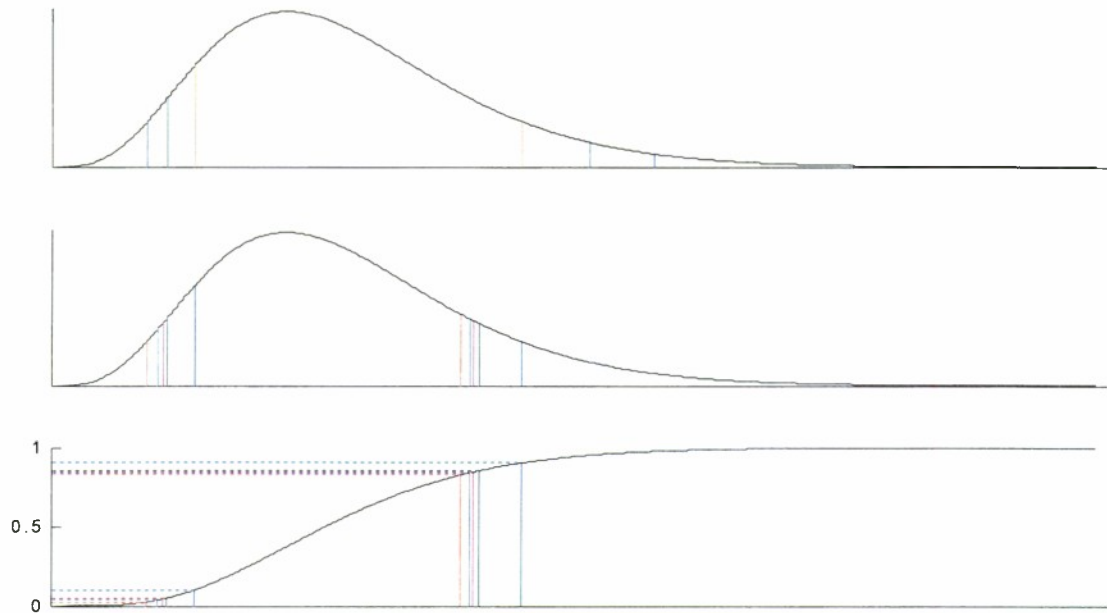


Figure 2. Rejection regions for a non symmetric distribution.

The two lower subplots illustrate an iterative approach to defining a rejection region that meets the design criteria. The initial step is obtained by applying the usual procedure as discussed above (blue lines). Looking at the density function shows the density higher at the left cutoff than at the right, indicating a need to shift both cutoffs to the left. This is done by decreasing the probability in the left tail and increasing the probability in the right tale by the same amount (green lines). This procedure is repeated until the densities at the two cutoffs agree to the desired level of accuracy.

Table 1. Iterative Procedure for choosing the rejection region.

Color	Left Boundary		Right Boundary	
	Prob. in tail	Density	Prob. in tail	Density
Blue	0.1	0.1281	0.1	0.0574
Green	0.05	0.0875	0.15	0.0811
Red	0.025	0.0571	0.175	0.0921
Cyan	0.0375	0.0736	0.1625	0.0867
Magenta	0.04375	0.0808	0.15625	0.0839

Table 1 shows the probabilities in the tails, the densities at the cutoffs and the colors used in the plot. Although this method does produce a rejection region meeting the desired criteria, it is labor intensive and requires the access to both the CDF and PDF functions.

3.3 The accumulate-toward-the-mode approach. Figure 3 illustrates how information from both the cumulative distribution function and the probability density function is used to define the accumulate-toward-the-mode distribution function (AMDF).

As in section 3.1, the rejection region is defined as the set of points where the PDF lies on or below a horizontal line as shown in the center subplot. This region consists of two tails, shown in yellow. The vertical lines indicate the boundaries of the left and right tails. The area of the left tail is the value of the CDF (upper subplot) at the left boundary point. The area of the right tail is one minus the value of the CDF at the right boundary point. The sum of these two areas is the significance level for this rejection region. This sum is also used to define the value of the AMDF (bottom subplot) at both the left and the right boundary points. As the horizontal line sweeps upward from zero to one, the area of both tails increases and their boundaries move closer together until the line reaches the peak of the PDF at its mode. Thus, the $AMDF(x)$ increases as x moves toward the mode and reaches its peak value of one when x is equal to the mode.

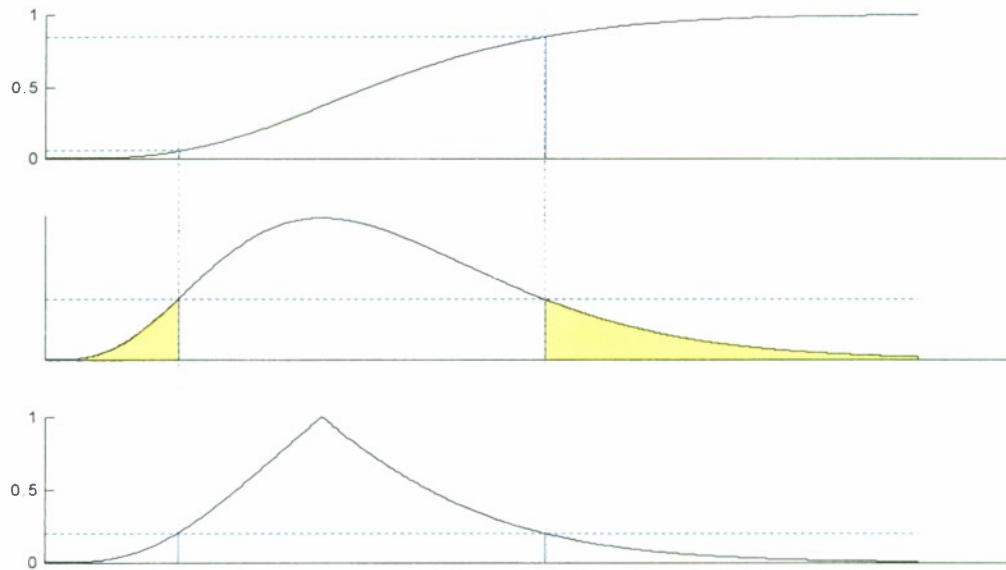


Figure 3. The accumulate-toward-the-mode distribution function.

The AMDF can be used to find a rejection region for significance level α by looking at a horizontal line at the height α in the lower subplot. The rejection region is the set of points where the AMDF lies at or below that line. The boundary points of the rejection region are the points where the AMDF crosses the line. Because of the way the AMDF was constructed, the probability density in the rejection region will always be less than that in the acceptance region, as desired.

4. Binomial distributions.

If a random variable Y is distributed $\text{bin}(n, p)$, then Y is the number of times an event, E , occurred in n independent trials and p is the probability of occurrence of E in each trial. The mean and standard deviation of Y are $\mu = n \cdot p$ and $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$.

By the Central Limit Theorem, we know that, for sufficiently large n , Y is distributed approximately $N\left(n \cdot p, \sqrt{n \cdot p \cdot (1 - p)}\right)$. In practice, this is usually restated in terms of Y/n , a point estimator for p , which is distributed approximately $N\left(p, \sqrt{p \cdot (1 - p)/n}\right)$.

4.1 Confidence intervals. There are two commonly used approaches for finding confidence intervals when working with binomially distributed data: using the normal approximation and using a table of the binomial distribution (i.e. the CDF). We will discuss each of these briefly, before looking at the development and application of a third approach, using the AMDF for binomial data.

Regardless of what method is used, one will not, in general, obtain a confidence interval with exactly the intended confidence level. This is a consequence of the fact that the binomial distribution is discrete.

4.1.1 The normal approximation. Probably the most common approach to generating confidence intervals for p is to use the normal approximation. To be more precise, the procedure generally used is to estimate the mean and standard deviation (μ and σ) from the data and then use a procedure designed for normally distributed data with unknown mean but known standard deviation. A slight variation is to use a procedure designed for normally distributed data where both the mean and the standard deviation are unknown, but the standard deviation is assumed to be fixed (i.e., although the mean may change in response to some “treatment”, the standard deviation will not).

The Central Limit Theorem is the justification for using the normal distribution in this way. This theorem is only valid when the sample size is “sufficiently large”, but gives no guidance for determining how large that might be. In practice, the necessary sample size depends on the true (but unknown) value of p , as well as the required accuracy.

Except for the case $p = 0.5$, the binomial distribution is not symmetric. This can affect both the rejection region and the confidence interval. The normal distribution will never show this non symmetry, regardless of sample size. The fact that, for the binomial distribution, σ varies with p also contributes to the asymmetry of binomial confidence intervals. The methods described above do not reflect this source of asymmetry either. There is a way to modify these methods to at least partially account for this factor, but it is seldom if ever used.

There is a rule-of-thumb that is sometimes used to decide if the sample size is big enough. This rule states that the value of p should be at least three standard deviations away from both zero and one. Since the value of p is unknown, this rule cannot be used to determine the sample size before the data are collected. However it can be used after the fact to indicate when the sample was too small for the estimated value of p . The sample size determined by this rule-of-thumb will be quite large for p near zero or one. As a result, one may be led to use a large sample

size in order to justify the methodology even though a smaller sample size would provide the desired accuracy.

4.1.2 Using the binomial CDF. The binomial CDF can be used to find rejection regions and confidence intervals using the same basic procedure described in section 3.1. Some adjustment must be made because of the discrete nature of the binomial distribution. The exact significance or confidence level may not be achievable. It can only be approximated. Similarly, the probabilities in the two tails will not, in general, be equal, only approximately so. Most tables cover only a limited range of sample sizes, typically up to 20. This procedure can easily be automated, which effectively eliminates the sample size restriction.

4.1.3 The binomial AMDF. Developing and working with the AMDF is a little different for a discrete distribution. The AMDF will only have nonzero values at the discrete set of points for which the PDF is nonzero. For the binomial distribution, this is the set of possible binomial outcomes.

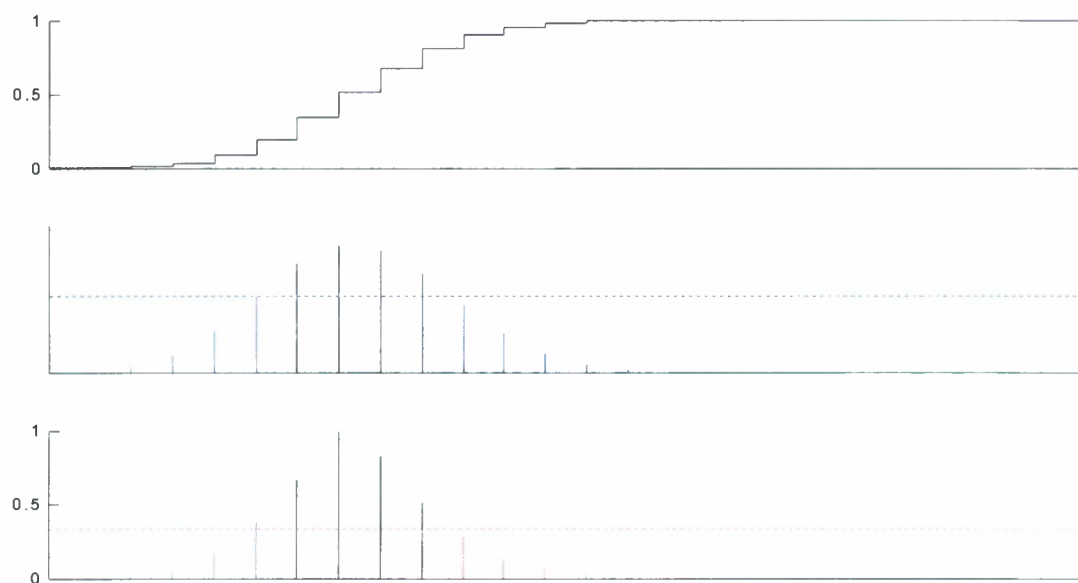


Figure 4. The binomial AMDF.

Figure 4 contains three subplots: the CDF at the top, the PDF in the center, and the AMDF. To determine the value of the AMDF for one of the possible binomial outcomes, draw a horizontal line through the value of the PDF at that point. The blue dashed line in the center subplot of Figure 4 is an example. The heights of the vertical lines in this subplot represent the values of the PDF function for all possible binomial outcomes. (Notice that the vertical scale in the center subplot does not match that of the other two subplots.) The sum of the values that fall at or below the dashed line (i.e., the vertical lines shown in blue) is the value of the AMDF at the point in question. This sum is the height of the vertical blue line in the lower subplot. To find a rejection region, draw a horizontal line at the height corresponding to the desired significance level. The magenta line in the lower subplot of Figure 4 is an example. All the possible outcomes for which the value of the AMDF falls on or below this line are in the rejection region. In the figure, these are also shown in magenta. When, as in this example, none of the AMDF values is

equal to the desired significance level, the intended significance level cannot be achieved exactly. The true significance level is somewhat smaller.

The binomial distribution shown in Figure 4 is a specific example with a specific sample size (25) and probability (0.3). For any fixed sample size, there is a family of PDF functions, one for each probability in the interval $[0,1]$. Recall that to determine a confidence interval for p , one must consider hypothesis tests for all of possible values of p . To do so with the accumulate-toward-the-mode approach, one will need the AMDF for each p in the interval $[0,1]$. Figure 5 depicts this family of AMDF functions for a binomial distribution with a sample size of 10.

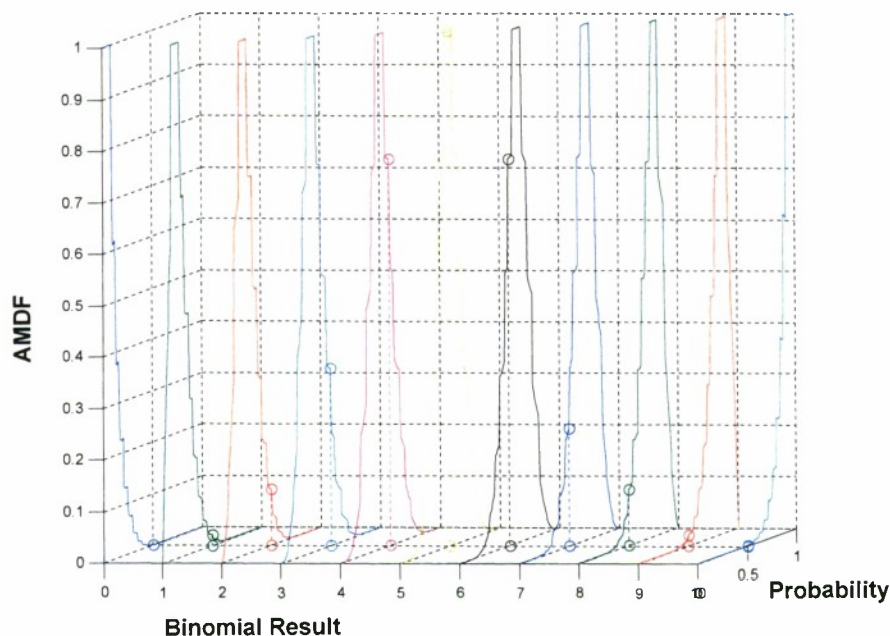


Figure 5. Family of binomial AMDF functions for a sample size of 10.

In this figure, the Binomial Result axis represents the possible outcomes, namely the discrete set of integers ranging from 0 to 10. The Probability axis represents the possible values of the probability of occurrence of the binomial event, all values in $[0,1]$. That is why the figure consists of a discrete set of curves, one for each binomial result. Each of these curves spans the full range, from 0 to 1, in probability. The individual curves are shown in different colors so they can be more easily distinguished. The height of one of these curves at any particular point indicates the value of the AMDF function for that point (i.e., binomial result and probability).

The plot of the AMDF function for a particular probability, similar to that in the lower subplot of Figure 4, is included in Figure 5. In this case, the binomial probability associated with the curve is 0.5. The sample size is, of course, 10. The values of this AMDF function are shown by vertical dashed lines with a circle at each end. As expected, these values are small for binomial results near the extremes (0 and 10) and reach a peak value of one at the binomial result of 5. Smaller values of the binomial probability lead to increasing values of the AMDF to the left side of the figure (smaller binomial results) and to decreasing values to the right side of the figure

(larger binomial results). Larger values of the binomial probability, of course, have the opposite effect. This is most easily seen for the most extreme cases (binomial results of 0 and 10).

For a single binomial result, the value of the AMDF function varies as the probability ranges from 0 to 1.0, tracing out the curve shown for that particular outcome. In each case, this value reaches its maximum value of 1.0 in an interval containing the binomial probability, p for which the binomial result equals the product of the sample size (10 in this case) and p .

The AMDF function can be used to find a rejection region for a hypothesis test or to find a confidence interval. These two tasks can be accomplished by focusing on vertical cross sections of the AMDF function, as depicted in Figure 5. Consider, for example, a hypothesis test with $p = 0.5$ as the null hypothesis. We have already seen that the intersection of the AMDF in Figure 5 with the plane, *Probability* = 0.5, is the binomial AMDF function for sample size 10 and probability 0.5. As in the discussion of Figure 4, the rejection region for a hypothesis test of this kind is the set of points (Binomial Results) for which the value of the AMDF is at or below the intended significance level.

To find a confidence interval, we focus on a different vertical cross section. For example, if the observed binomial outcome is 7, we would look at the intersection with the plane, *Binomial Result* = 7. Values of the AMDF function in this plane are shown in blue in Figure 5. The confidence interval consists of all probabilities for which this curve lies above $(1 - C/100)$, where C is the intended confidence level (expressed as a percentage).

4.2 Confidence regions for multivariate distributions.

In section 2, we discussed the close relationship between hypothesis tests for the value of a parameter and confidence intervals for the value of that parameter. These ideas easily extend to the multivariate case. Consider a vector parameter instead of a scalar for the hypothesis test and a region of a multidimensional space which can be expected, with the prescribed level of confidence, to contain the true value of that vector parameter. Section 4.2.1 illustrates this approach for a bivariate binomial distribution when the two components are independent. The extension to higher dimensions is straightforward. If the assumption of independence is removed, computation of the multidimensional PDF is more complicated, but if that can be accomplished, the same procedure will work. Of course, visualization of the AMDF function becomes more difficult. Even for the bivariate case, the analogs of Figures 4 and 5 would require 3 and 5 dimensions respectively.

The idea of confidence regions for multivariate distributions is not new. For multivariate normal distributions, confidence ellipsoids, based on Hotelling's T^2 test, can be generated. This is a generalization of the one dimensional procedure for the normal distribution that was discussed earlier. One could certainly apply it to the multivariate binomial case in much the same way the normal approximation is often used for the binomial distribution. This procedure would have shortcomings similar to those pointed out for the univariate case.

There is a multivariate extension of the CDF function, but it is not useful for the generation of confidence regions. Applying the AMDF approach to the multivariate binomial distribution allows one to generate confidence regions based on the true distribution. Regardless of what method is used, one will not, in general, obtain a confidence region with exactly the intended confidence level. This is a consequence of the fact that the binomial distribution is discrete.

4.2.1 The AMDF for a bivariate binomial distribution. Computing and using the AMDF for the bivariate binomial involves basically the same process that was used above for the binomial distribution. Figure 6 contains two subplots. The upper subplot shows the PDF of a bivariate binomial distribution, (x, y) , with x and y distributed independently. The lower subplot shows the corresponding AMDF. Sample sizes and probabilities for x and y are as shown. The x and y axes, representing the possible outcomes, have been normalized by dividing by the sample size. Thus, the possible values in the x direction range from zero to one in increments of 0.2. In the y direction, they range from zero to one in increments of 0.1. The color coding in the two plots is intended to aid visualization of two processes, calculating the AMDF and using the AMDF to define a rejection region. The focus for this discussion is the outcome $(x, y) = (1, 4)$, corresponding to the point $p_0 = (0.2, 0.4)$ in each subplot. In the upper plot, the probability of

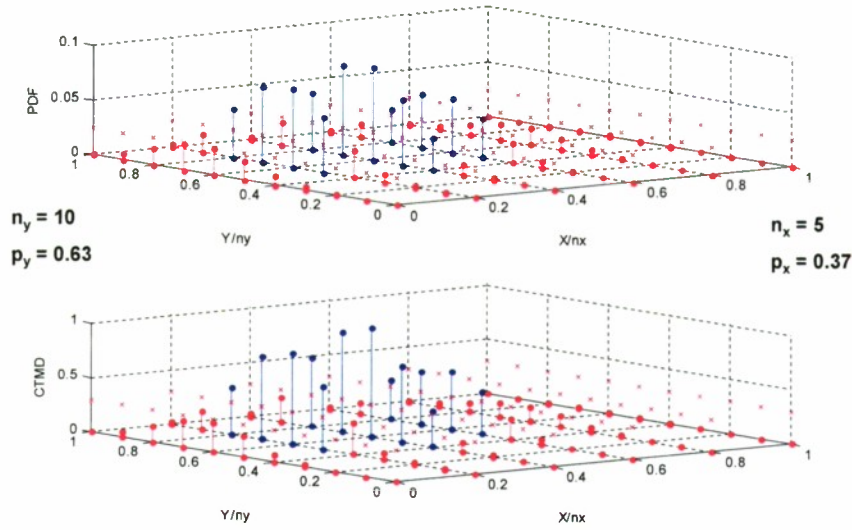


Figure 6. An example of the bivariate binomial AMDF.

occurrence of each possible outcome is indicated by a vertical line with a solid dot at each end. (i.e., The length of each vertical line is the probability of occurrence for that outcome, so the upper dot is at that height.) Where probability of a given outcome is less than or equal to the probability of occurrence of p_0 , the vertical lines are colored magenta. The blue lines correspond to points whose probability is greater than that of p_0 . The magenta x's are plotted at a height corresponding to the probability of p_0 above each possible outcome.

The value of the AMDF at p_0 is the sum of the probabilities of all the outcomes whose probability is less than or equal to the probability of occurrence of p_0 (i.e., the sum of the lengths of all the magenta vertical lines in the upper subplot). This is the length of the vertical line at p_0 in the lower plot. (Notice the difference in the vertical scales in the two subplots.) A similar calculation for each possible outcome (i.e., for each possible combination of x and y) leads to the AMDF function as shown in the lower plot. In this plot also, the magenta x's indicate the height of the function at p_0 and the vertical lines that do not reach above that level are shown in magenta. These points make up a rejection region for a hypothesis test with null hypothesis: $p_x = 0.37$ and $p_y = 0.63$; and significance level, α , equal to the value of the AMDF function at p_0 . The

vertical blue lines correspond to possible outcomes in the acceptance region. These are the possible outcomes for which the point $(0.37, 0.63)$ would be in a $100 \cdot (1 - \alpha)$ confidence region.

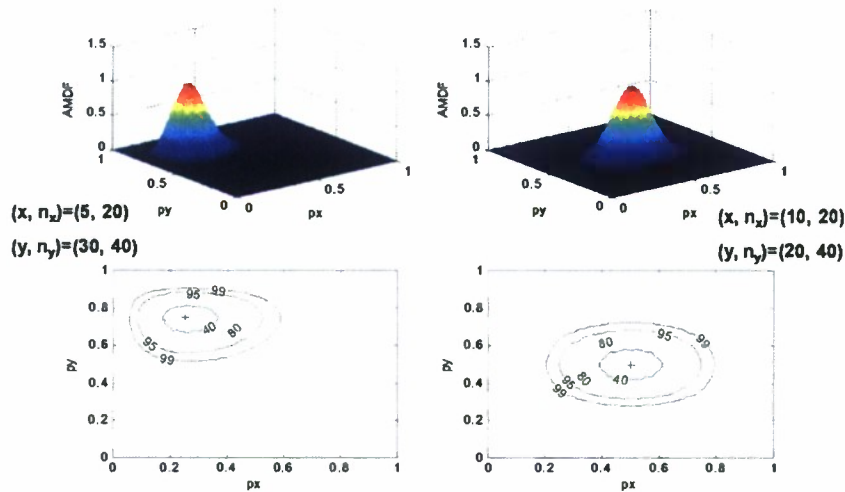


Figure 7. Confidence regions for bivariate binomial distributions.

Figure 7 illustrates how confidence regions are obtained from the AMDF. The upper subplots show the AMDF, for two possible outcomes from a bivariate binomial distribution. The outcomes and sample sizes for x and y are as shown. In either of these subplots, a confidence region for a given confidence level, C , would be the set of points, (p_x, p_y) , where the value of the AMDF is greater than $(1 - C/100)$. The boundaries of several such confidence intervals are shown in the lower subplots. The '+' in these subplots shows the point estimate for (p_x, p_y) . The confidence regions all contain the point estimate, but they are not all symmetric. In fact, the lower right subplot shows the only case for which the binomial confidence intervals are symmetric, namely when the point estimate for (p_x, p_y) is $(0.5, 0.5)$. In both subplots, the confidence regions are elongated. This is due to the difference in sample sizes. The estimate of p_x is less precise because the sample size is smaller.

5. An Application.

In a recent study the Army Materiel Systems Analysis Activity (AMSAA) examined the level of protection provided by two different helmet designs. The Integrated Casualty Estimation Model (ICEM) was used to evaluate the effectiveness of each helmet. This resulted in two binomially distributed statistics. For reasons that will be explained below, these data were subjected to two separate analyses. Only one of these was, in the end, actually used for the study. However, both are discussed here, because they demonstrate two different applications of accumulate-toward-the-mode methods with binomially distributed data.

5.1 The original question. The original question regarding analysis of the helmet study data was, ‘Given two binomially distributed statistics, X and Y , both from a sample size of 500, how far apart do X and Y have to be to show a statistically significant difference?’

To address this question, we consider a hypothesis test with null hypothesis, $p_x = p_y$. One way to conduct this test would be to generate a confidence region for (p_x, p_y) , based on the observed outcome (X, Y) . If the line $p_x = p_y$ does not intersect this region, then the null hypothesis is rejected.

Generating a set of AMDF tables for problems of this kind would require a significant effort. Furthermore, searching through such a set of tables to determine the boundary of a confidence region would be tedious and time consuming at best. Fortunately, this is not necessary. It is not difficult to develop an automated procedure to compute the value of the AMDF for a given bivariate binomial outcome, (X, Y) , and a hypothesized pair, (p_x, p_y) , of probabilities.

A single execution of this procedure would be needed to test the hypothesis that the probability pair, (p_x, p_y) , is the pair of probabilities associated with the bivariate binomial distribution that produced the binomial outcome, (X, Y) . Multiple executions could be used to find a confidence region, or to test a compound hypothesis test such as $p_x = p_y$. This would answer the question of whether, for a specific outcome (X, Y) , X and Y are statistically different. Another layer of repetition, spanning the possible bivariate binomial outcomes, could address the broader question above, “How far apart do X and Y have to be to be significantly different?” This would be computationally intensive, but certainly doable. A couple of observations lead to a quicker answer to this broader question.

It is not necessary to find the entire confidence region: just its intersection, if any, with the line $p_x = p_y$. Since a confidence region with confidence level C is the set of points, (p_x, p_y) , for which $\text{AMDF}(X, Y)$ exceeds $(1 - C/100)$, it is sufficient to look at the maximum of $\text{AMDF}(X, Y)$ on the line.

It is not difficult to show that the probability, p_m , that maximizes the $\text{PDF}(X, Y)$ along the line $p_x = p_y$ is the average of the point estimates for p_x and p_y . (i.e., $p_m = (X + Y)/1000$ in our current example. Considering the way the AMDF is related to the PDF, one might expect that the maximum of the AMDF on the line $p_x = p_y$ would also fall at or near the point (p_m, p_m) . Figure 8 was generated with the assumption that this expectation is met. The AMDF values

shown in the figure are the values for the point (p_m, p_m) . The actual value of the AMDF could be greater than this, but not smaller.

Each curve in this figure represents the maximum values of the AMDF on the line $p_x = p_y$ for (X, Y) pairs with a given absolute difference. The difference corresponding to each curve is indicated in the legend.

Where the curve falls above the horizontal dashed line at the 0.05 level, the null hypothesis, $p_x = p_y$, cannot be rejected for a 0.05 significance level. Where the curve falls at or below that line, the hypothesis can be rejected for that significance level. From Figure 8, we can conclude that the ability to distinguish between binomial populations depends not only on the sample size and separation, but on where within the possible range of values X and Y fall. For our example, the difference would have to be around 40 or greater to guarantee statistical significance, at the 0.05 level, across the full range of possible outcomes.

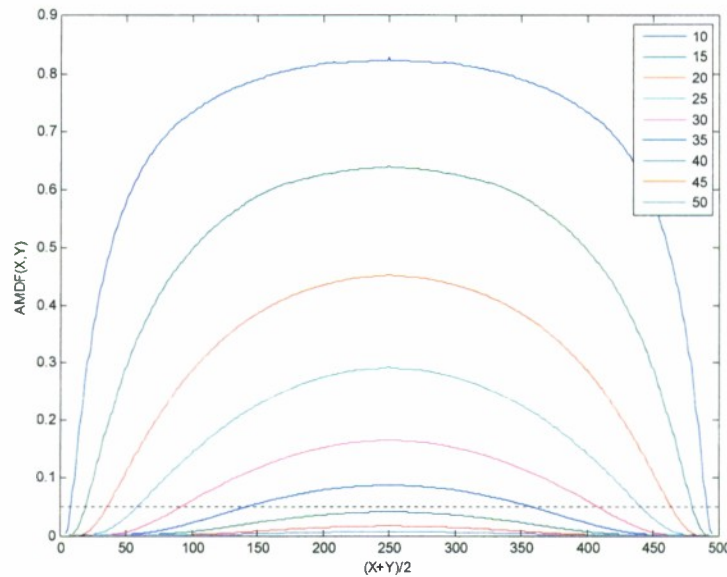


Figure 8. Maximum values of $AMDF(X, Y)$ on the line $p_x = p_y$ for various absolute differences $(|X - Y|)$.

More detailed calculations at a number of points on several of the curves in Figure 8 showed that the maximum of the AMDF on the line $p_x = p_y$ does occur at a probability of p_m (or at least within 0.01 of that value). This may be due to the fact that the sample sizes are equal. Other cases have yet to be investigated.

5.2 The amended question. The analysis discussed in the previous section is a good example of the AMDF approach applied to a bivariate binomial distribution. However, it was not the appropriate analysis for the helmet study. This study was one half of a tradeoff evaluation. One of the helmet designs was obtained from the other by removing a small area around each ear to improve hearing, and therefore situation awareness. The object of the helmet study was to determine the cost, in terms of increased vulnerability, of this change. Thus, it was necessary to determine, from a vulnerability standpoint, if the two designs were different and, if so, to quantify the difference. However, the analysis of the previous section was based on the wrong statistic.

The binomial statistics, X and Y , were generated by a Monte Carlo simulation. In each replication of the simulation, a model helmet was tested against a random representation of the threat. However, as a variance reduction technique a single set of 500 threat representations (fragmentation patterns) was generated and each of these was used for two simulation runs, once against each of the helmet designs. (i.e., once for X and once for Y) Clearly, X and Y are not independent, as assumed in the previous section. Correct analysis of this data must account for the pairing of the two data sets. Typically, this is done by looking at the differences between the paired data. In this case, that leads to the definition of a new statistic, $Z = Y - X$. We would not usually expect Z to be binomially distributed. However, since one design was obtained from the other by removing a small area on either side, it is possible to get a worse injury with the new design, but not the other way around. A pair of Monte Carlo replications with a given threat representation could have only three results: no injury in either case, an injury for the new design only (represented by the Y statistic), or an injury for both designs. Thus, Z is binomial with a sample size of 500. Given an outcome for Z , we can use the AMDF to find a confidence interval for p_z , the difference between p_y and p_x . If zero is not in that interval, not only can we conclude that the two distributions are different, but we have an interval estimate for the difference between their underlying probabilities.

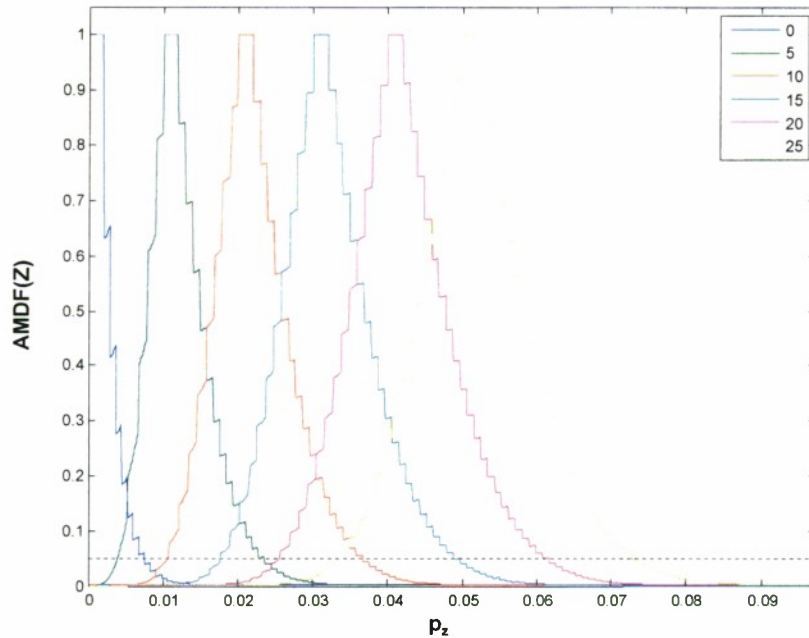


Figure 9. AMDF(Z) for various outcomes.

Figure 9 shows the AMDF function for several possible outcomes ranging from 0 to 25. For each of these outcomes, the corresponding 95% confidence interval is the region where the curve for that outcome lies above the horizontal dotted line at the 0.05 level. Except for the case when the outcome is 0, these confidence intervals do not include zero. Plots for outcomes between zero and five are not shown here, but those curves are much like the one for an outcome of five compressed toward the left side of the figure. Therefore, for all non-zero outcomes, the hypothesis, $p_x = p_y$, can be rejected. In addition, a confidence interval for the difference

between the two probabilities is obtained. This is a much stronger result than was produced by the original analysis. In this case, rejection of the null hypothesis does not depend on the sizes of or the absolute difference between X and Y , only on the fact that they are different.

5.3 Comparing methods. Figure 10, shows confidence intervals for a few of the possible outcomes in the helmet study. The outcomes actually observed in the study were in the range of outcomes shown in this figure. The confidence intervals in red were obtained by use of the normal approximation. Those in blue were generated using the AMDF. In the figure, the ‘x’, and ‘+’ indicate the lower and upper limits of the confidence interval, while the open circle indicates the point estimate.

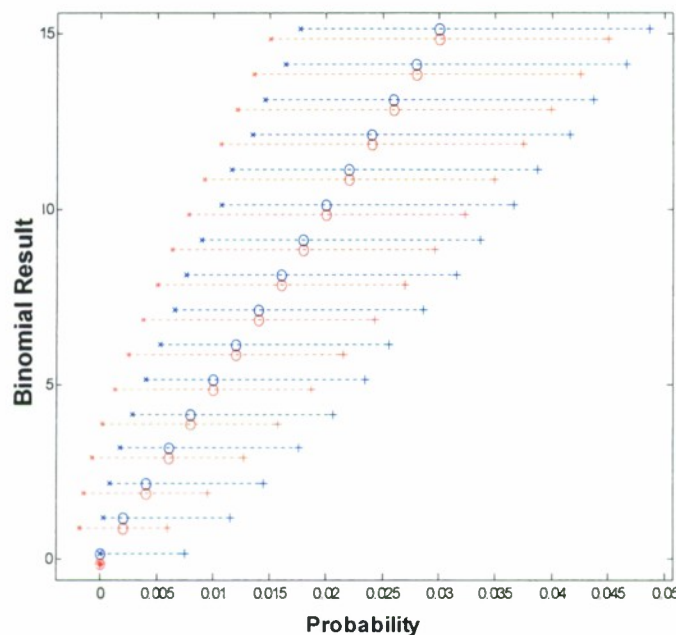


Figure 10. Comparison of Binomial and Normal 95% confidence intervals for the helmet study.

Two points that were mentioned earlier are visible here. The confidence intervals generated by the normal approximation are symmetric about the point estimate, and negative values are included in the confidence intervals for a few of the smaller outcomes. The second point leads to an incorrect result for those few outcomes: since zero lies in the confidence interval, the null hypothesis cannot be rejected. For these cases, the interpretation of the helmet study results would be reversed. A third point, not mentioned before is that the “normal” confidence interval collapses to a single point when zero is the binomial outcome.

When working with discrete distributions, one cannot, in general, get precisely the confidence level (or, for hypothesis tests, significance level) intended. This is reflected in the vertical jumps of the AMDF function. (See Figure 9) If the horizontal line in the figure crosses the AMDF curve at a vertical jump, then the intended confidence level cannot be achieved. However, when the AMDF, is used as described above, the true confidence level will always be greater than or equal to the intended level. When the normal approximation is used, there may be additional deviation from the intended confidence level related to the accuracy of the approximation and the symmetry of the distribution.

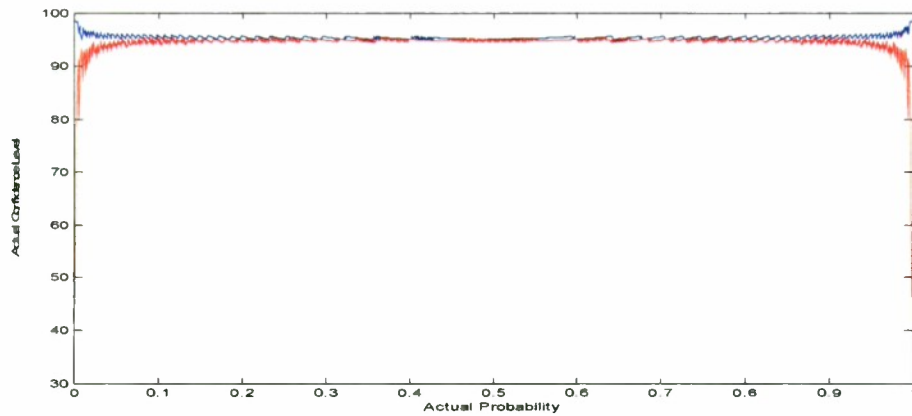


Figure 11. Actual confidence levels.

Figure 11 shows the confidence level achieved when applying the normal approximation and AMDF to the helmet study results. Recall that the sample size was 500 and the intended confidence level was 95%. The actual confidence level is a function of the true binomial probability (i.e. the actual probability that the binomial event will occur in any given trial). For each possible probability of occurrence, p , the red line shows the confidence (i.e., 100 times the probability) that the interval generated using the normal approximation will, in fact, contain p . Similarly, the blue line shows the actual confidence when the AMDF is used. With this large sample size, the confidence levels in the center of the figure, though somewhat noisy, are fairly close to the intended level of 95%. Near the sides of the figure, the blue curve trends up, towards 100%, while the red trends down toward 0% before jumping to 100% for actual probabilities of 0 and 1.

Similar curves for a sample size of 50 are shown in Figure 12. Once again, the intended confidence level is 95%. In this figure, more detail is visible, but the trends are basically the same, although the variability is greater and extends further into the center of the figure. With the AMDF, the actual confidence level is always at or above the design level. When the normal approximation is used, the actual confidence level is consistently low, dramatically so for probabilities near (but not equal to) zero and one. The actual confidence levels continue to degrade as the sample size decreases.

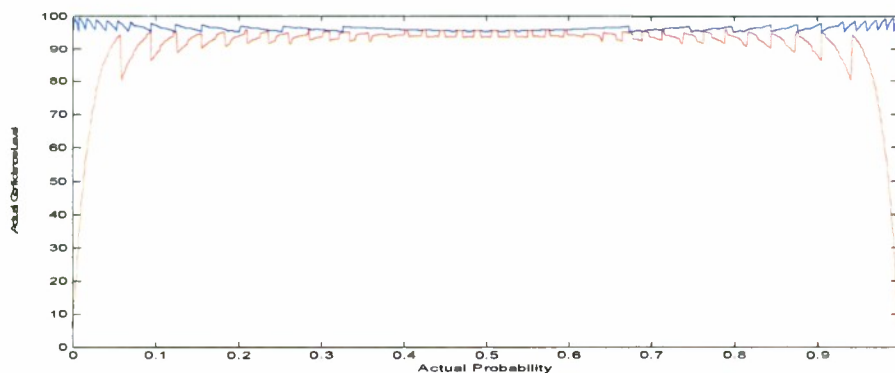


Figure 12. Actual confidence levels.

6. Observations and Conclusions.

An analyst, working with binomially distributed data cannot go to a table of AMDF values to find the rejection region for a hypothesis test or to obtain an interval estimate of a binomial probability. Such tables do not exist. However, it is relatively easy to develop software to provide the needed information. Once this software is available, it is easier and quicker to do the analysis than it would be with a table. Appendix A contains a listing of a MATLAB[®] routine for this purpose. Software such as this makes the AMDF a useful alternative to either the binomial CDF or the normal approximation for analysis of binomial data. There are a number of reasons why the AMDF approach is preferable. Use of the AMDF, as described above, will give the smallest confidence interval with at least the desired level of confidence. This is true regardless of sample size: there is no need to use a larger sample size to justify the analysis method. Sample size can be chosen on the basis of required accuracy, and the AMDF can be used, before the experiment, to make that determination. Finally, the AMDF also works well for multivariate binomial data.

The helmet study provided an opportunity to demonstrate the use of the AMDF approach. Several other useful lessons were also learned. It is important to understand the problem and fit the analysis to that problem. Use variance reduction techniques when they are applicable. If the analysis and the data collection are planned in advance, that planning can include a preliminary analysis to determine the sample size.

REFERENCES

1. Hogg, R. V., and Craig, A.T., Introduction to Mathematical Statistics (Fourth Edition), Macmillan Publishing Co., Inc., New York, 1978.

THIS PAGE INTENTIONALLY LEFT BLANK.

APPENDIX A – MATLAB® Implementation

The MATLAB[®] code below computes a confidence interval based on the AMDF for binomially distributed data. The first routine, BinCI determines the confidence interval for the given sample size, confidence level, and binomial result. The forth input, probabilityStep allows the user to trade of accuracy vs. computation time. The probabilityStep should be small to ensure accuracy; 0.001 is probably adequate for most applications, although 0.0001 was needed for some of the plots in this paper in order to show the detailed structure of the AMDF.

The second routine, binocv, returns the value of the AMDF function for a range of possible binomial probabilities. It makes use of a third function, binopdf, which returns values of the binomial probability density function. A listing of binopdf is not provided because it is a MATLAB[®] function included in the Statistics Toolbox. Implementation in other languages is not difficult, although it may require an implementation of the binomial probability density function.

```
function [ ci ] = BinCI( outcome, sampleSize, confLevel, probabilityStep )
%BINCI returns an AMDF confidence interval for the actual probability
%   of success of a binomial distribution
%   input data
%       sampleSize = N, the number of independent trials
%       outcome = n, the test result (i.e., number of successes that occurred
%           in N trials)
%       confLevel, the objective confidence level
%       probabilityStep, the probability step size
%       Note: determines the accuracy to which the endpoints of the
%           confidence interval are determined
%
%   intermediate values
%       a - significance level for the family of hypothesis tests
%       p - vector of hypothesized probability-of-event values for one trial
%
%   output data
%       ci - returned value, containing
%           n
%           N
%           lower bound of the confidence interval
%           point estimate for the actual probability of success
%           upper bound of the confidence interval
%           objective confidence level
%           actual confidence level (for this outcome and sample size)
%
N = sampleSize;
n = outcome;
coLev = confLevel;
step = probabilityStep;

a = 1 - (coLev / 100.0);
p = (0:step:1)';
ci = zeros(1,7);

x = binocv(n, N, p);
L = x > a;
y = [p(L) x(L)];
lc = y(1,1);
```

```

uc = y(end,1);
m = n/N;
tcl = 100*(1.0 - max(x(~L)));
ci = [n N lc m uc coLev tcl ];
if(isdeployed)
    fprintf(1,'%10d%10d%10.3f%10.3f%10.3f%10d%10.3f\n', n, N, lc, m, uc,
coLev, tcl);
end
end

function [ cv ] = binocv(n, N, p )
%BINOCV Compute the critical value, based on the AMDF, for a particular
% outcome as a function of p
% data
%     N - number of independent trials
%     n - test result (i.e., number of successes that occurred in N
%         trials)
%     p - vector of hypothesized probability-of-event values for one
%         trial
% output
%     cv - a vector of critical values, one for each probability in p
%

lp = length(p);
cv = 0.0 .* p;
for i = 1:lp;
    mat = binopdf((0:N), N, p(i));
    lv = mat <= mat(n+1);
    cv(i) = sum(mat(lv));
end

```

binopdf(x, N, p) is the binomial probability density function. The inputs are the binomial outcome, x, the sample size, N, and the probability-of-event in each trial.

THIS PAGE INTENTIONALLY LEFT BLANK.

APPENDIX B - DISTRIBUTION LIST

APPENDIX B - DISTRIBUTION LIST

No. of Copies	Organization
15	Director US Army Materiel Systems Analysis Activity AMXAA-B/Ms. Brenda Platt AMXAA-W/Mr. Robert Chandler AMXAA-WS/Ms. Regina Kistner/Dr. Dwayne Nuzman/ Mr. Matthew Rosenblatt AMXAA-WM/Mr. John DeVido/Mr. Gary Comstock AMXAA-SI/Mr. John Mazz AMXAA-LR/Mr. John Nierwinski/Mr. Martin Wayne 392 Hopkins Road Aberdeen Proving Ground, MD 21005-5071
1	Director US Army Evaluation Center TEAE-RE/Dr. Michael Cushing 4120 Susquehanna Avenue Aberdeen Proving Ground, MD 21005-3013
1	US Army TRADOC Analysis Center ATRC-PR/Ms. Susan Matus 255 Sedgwick Avenue Fort Leavenworth, KS 66027-2345
1	US Army TRAC- White Sands Missile Range ATRC-F/COL Gary Lambert Bldg 1400 Martin Luther King Drive White Sands Missile Range, NM 88002-5502
1	Defense Technical Information Center DTIC-O 8725 John J. Kingman Road Fort Belvoir, VA 22060-6218
1	Office of the Deputy Chief of Staff, G-4 DALO-ZX/Mr. John Corcoran 500 Army Pentagon (Room 1E394) Washington, DC 20310-0500